SIDECAR:

A first step toward interoperable JCIDS docs.

*s*emantically *i*nformed *d*ynamic *e*ngineering of *c*apabilities *a*nd *r*equirements

End-of-Stage-1 Demo – 16 December, 2010

CYCORP

www.CYC.com

Doug Lenat
Ben Rode
Nathan Winant
Robert Kahlert
David Baxter
Steve Collins

Assisted by:

Everware-CBDI

Dave Mayo
Gary Berg-Cross

NPS

Chris Gunderson

Supported by:

HQMC Intelligence Dept

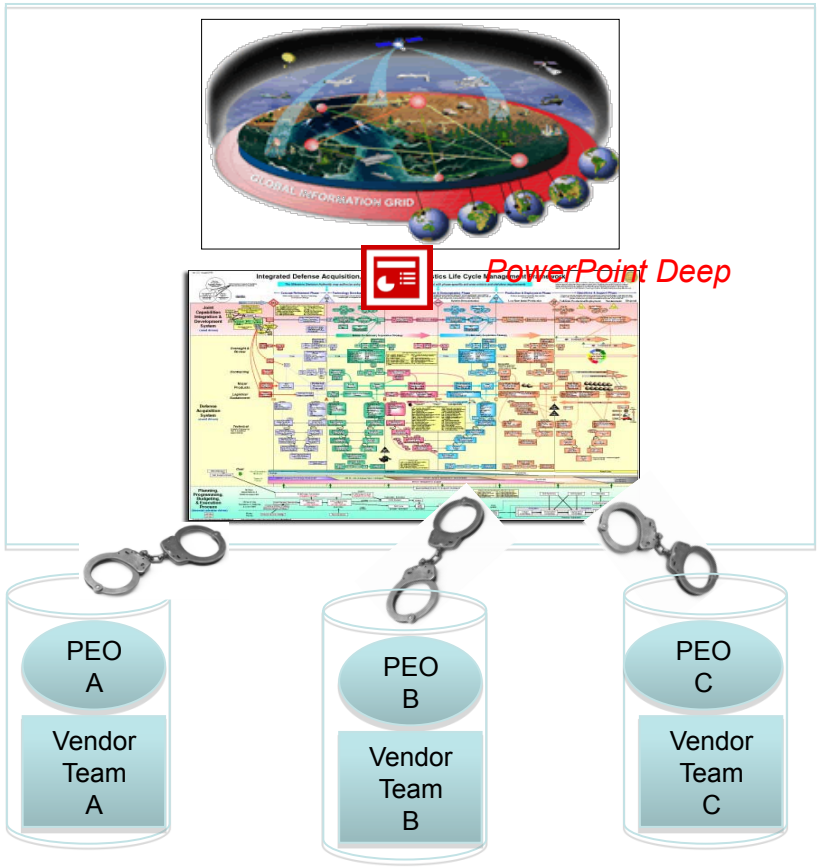
Intel Integration Division, Capabilities Development Directorate, DC CD&I through MTCSC

7718 Wood Hollow Drive, Suite 250, Austin, Texas 78731 Lenat@Cyc.com 512-342-4000

The Context: Enterprise Architecture

- **Key Challenge: Current Requirements and Acquisition Processes do NOT lend themselves to an Enterprise Architecture Approach.**

Policy, paperwork, and bureaucracy

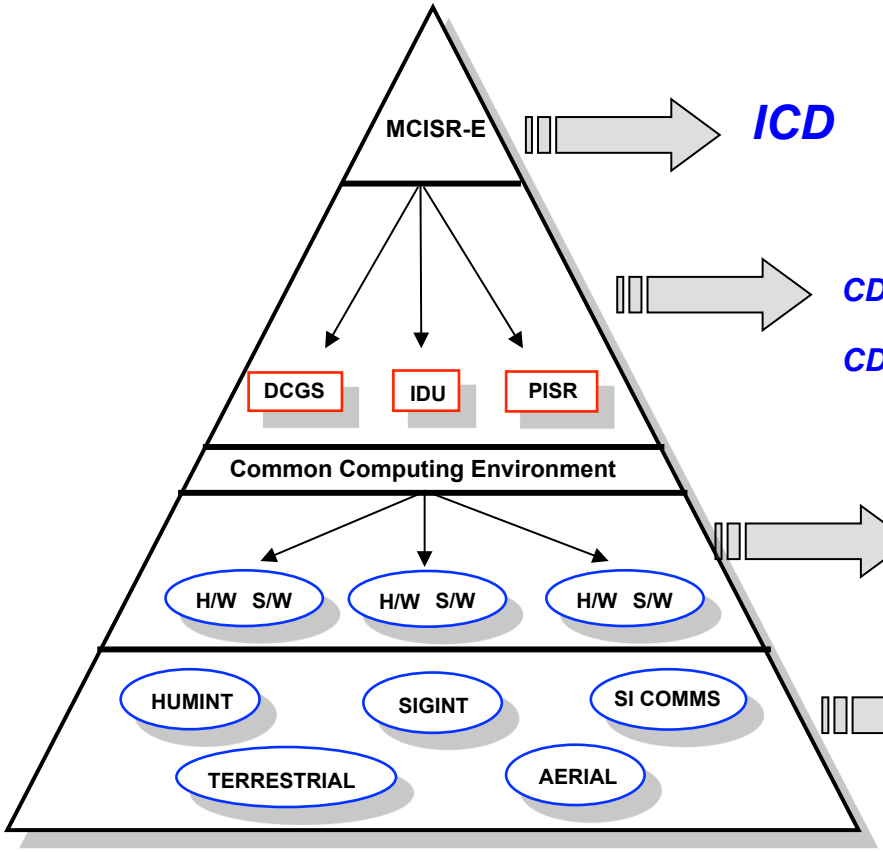


Programs deliver large increments of sub-optimum capability late and over budget

GOALS

- **Define and develop a "value-delivery chain."**
- **Accelerate intelligence data collaboration.**
- **Research, develop, and procure systems that complement one another.**
- Via 1,2,3: Deliver synoptic view of battlespace to commanders.
- "Operationalize" S&T efforts.
- Ensure "repeatability."

The Context (2): OPT Strategy for Enterprise-Level Capabilities Documents



ICD

CDDs that are exceptional because…

CDDs that are extraordinary because…

Common Computing Environment CPD(s)

Capability set CPDs

The Context (3): What is Different about CDDs?

- CDDs that are exceptional because…
  - **CDD will** not be **for a single increment of development or a single material solution (box), but is a Capability Description Document for an enterprise architecture (Concurrence for Beta-Test of Process Improvement required from J8).**
  - **Any number of diverse Capability Set increments will be defined in and be produced off the CDD baseline (Concurrence for Beta-Test of Process Improvement required from J8).**
- CDD's that are extraordinary because…
  - **Define an enterprise architecture and common computing environment (OV-1 MCISR-E ICD).**
  - **Adopt commercial best practices and use common components between PISR, IDU and DCGS.**
  - **Specify, within those CDDs, a business process.**
  - **Codify a developmental process for USMC Intel capabilities in a requirements document.**
  - **Are written in "mark-up language" (machine-reasonable) to facilitate integration.**
  - **Define an open system architecture with compliance required from vendors and developers.**

The Context (4): The problem SIDECAR is tackling

- Acq. & Req's. needs to model programs, ICDs, CDDs, CPDs, relevant policies, technologies,… so new capabilities/requirements doc's.
  - can be produced faster and with higher accuracy and precision and interoperability,
  - can be kept up to date,
  - can trigger alerts when things change that make parts of them stale or noncompliant,
  - can predictively analyze *possible* changes (i.e., infer the impact of that change on existing programs, given their requirements and interdependencies.)

- So what is the bottleneck to accomplishing this functionality?
  - S^3 (size, speed, security) ← often the *apparent* bottleneck
  - Lack of deep semantic understanding ← more daunting problem obscured by S^3
    - "understanding": represent the knowledge and reason with it (at least to human levels)
    - It would be reckless to delegate this 99% of this task to a group of brittle idiot savants.
    - And yet, in effect that is what is going on, today, since that is the level of intelligence in today's databases and textual document preparation and processing software tools.

The Bottom Line for Today

- You'll see a demo of a system (SIDECAR) in which:
  - The user (a req. officer) is writing a CDD
  - As the user "fills out" each part of each section of that doc.:
    - SIDECAR uses that newly entered information to infer something
      - Notice something out of compliance with existing policies and req's.
      - Constrain future "menu choices" (special case: 0 or 1 options remain)
    - SIDECAR reacts appropriately to that newly *inferred* conclusion
    - Analogue: TurboTax helping you fill out your 1040 tax return

- What we're asking from you, going forward:
  - Support **MCISR-E*** documentation strategy as a test case of this "active deep semantics" approach to crafting/updating capabilities requirements
  - Assistance in exposing this to the broader requirements community and in forging a partnership between VCJCS's **CDTM**** effort and SIDECAR

*** Marine Corps Intelligence, Surveillance, and Reconnaissance Enterprise**
**** Capability Development Tracking Management System (Faccina Global)**

TurboTax Deluxe 2009

File Edit View Tools Online Help

TurboTax Deluxe

Federal Refund $0

Forms Print Center Help Center

Personal Info | Federal Taxes | State Taxes | Wrap-Up | Print & File

Find Show Topic List

Wages & Income | Deductions & Credits | Other Tax Situations | Federal Review | Error Check

Enter 1099-MISC Information

Fill in **any amounts** that appear on your Form 1099-MISC, in the **corresponding boxes** below. (Enter one 1099-MISC at a time.)

Who Paid You?

Payer's Tax ID No:

Employer ID No. Or Soc. Sec. No.

Boxes 1-15b

1 - Rents

2 - Royalties

3 - Other Income

4 - Federal Tax Withheld

5 - Fishing Boat Proceeds

6 - Medical Payments

7 - Nonemployee Comp

8 - Substitute Payments

9 - N/A

10 - Crop Insurance

13 - Golden Parachute Pmts

14 - Gross Attorney's Fees

15a - Section 409A Deferrals

15b - Section 409A Income

Boxes 16-18

16 - State Tax State 18 - State

Flags

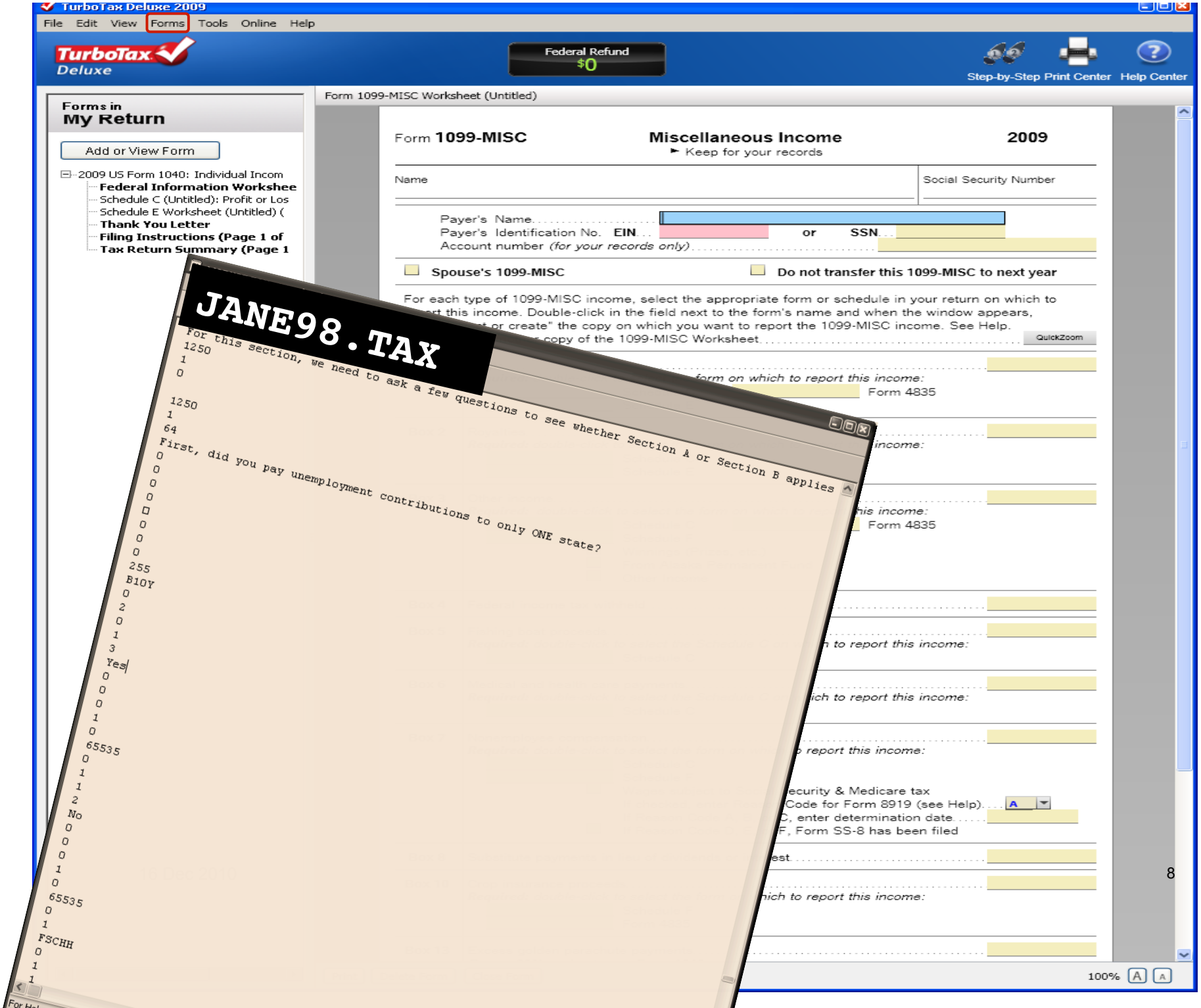
Flag this page to come back to it later

Add a New Flag

Live Community

Get help from other TurboTax customers

View Community

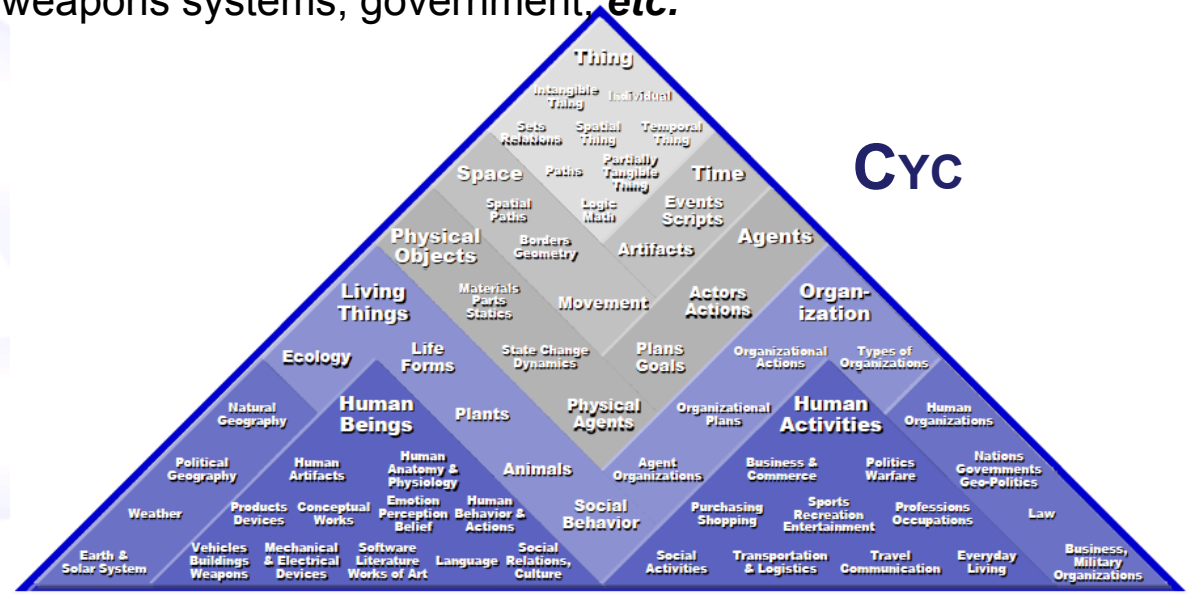

TurboTax Deluxe 2009
File Edit View Forms Tools Online Help
TurboTax Deluxe
Federal Refund
$0
Step-by-Step
Print Center
Help Center
Forms in My Return
Add or View Form
2009 US Form 1040: Individual Incom
Federal Information Workshee
Schedule C (Untitled): Profit or Los
Schedule E Worksheet (Untitled) (
Thank You Letter
Filing Instructions (Page 1 of
Tax Return Summary (Page 1
Form 1099-MISC Worksheet (Untitled)
Form 1099-MISC
Miscellaneous Income
Keep for your records
2009
Name
Social Security Number
Payer's Name
Payer's Identification No. EIN
or SSN
Account number (for your records only)
Spouse's 1099-MISC
Do not transfer this 1099-MISC to next year
For each type of 1099-MISC income, select the appropriate form or schedule in your return on which to
this income. Double-click in the field next to the form's name and when the window appears,
or create" the copy on which you want to report the 1099-MISC income. See Help.
copy of the 1099-MISC Worksheet
QuickZoom
Form 4835
Form 4835
100%
JANE98.TAX
For this section, we need to ask a few questions to see whether Section A or Section B applies
1250
1
0
1250
1
64
First, did you pay unemployment contributions to only ONE state?
255
B10Y
2
0
1
3
Yes
65535
No
65535
FSCHH

What does SIDECAR start with?

(1) a detailed understanding of CDDs

(2) a partial model of related efforts (eg MCISR-E) and organizations (eg JCB) and policies (eg Net-Ready Key Performance Param., DOD Enterprise Architecture)
- This has increasing returns on usage (as it runs on more and more documents)

(3) millions of pieces of real world knowledge and "common sense" – **microtheories** – about physical objects and processes, time, space, goals, containment, roles, causation, belief, intelligence, arguments, countries, climate, software, networks, computers, transportation, weapons systems, government, ***etc.***



How does SIDECAR represent all that?

Each statement is represented simultaneously in several different ways:

(1) *superficially:* record the specific blanks that are filled in, boxes checked, menu items chosen, etc., on the SIDECAR screens

(2) *formally:* record each piece of knowledge in higher order logic (n^{th}-order predicate calculus), so it is fully machine-reasonable
- **The user never sees this!**

(3) *"semantically"* (as in Semantic Web): node & link graphs/diagrams such as the CONOPS visualizer
- A display generated from a subset of the formal assertions

(4) Hyperlinked English prose: the growing CDD document
- Automatically-generated glosses of a subset of the formal assertions
- Links point back to the assertions and, through them, the entry screens

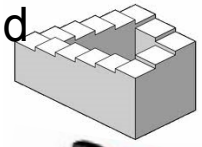
(5) the flattened English-text-only CDD document
- Remove the hyperlinks from (4)

So what does it actually *do*, with that knowledge?

As each new assertion (about the target system, or about the CDD) is added, SIDECAR runs an inference engine to conclude new things it believes to be true, based on everything that it has been told so far.

- ⌉ **And then it reacts appropriately to the newly inferred assertions.**

– Sometimes it deduces a contradiction with what it was already told



- **Ask the user to resolve that, or explicitly defer resolving it**

– Sometimes it infers some constrained choices for menus/options.

- **Special case:** the number of options goes to **0**: **Eliminate that question entirely**
- **Special case:** the number of options goes to **1**: ***Guess* that as the correct value**

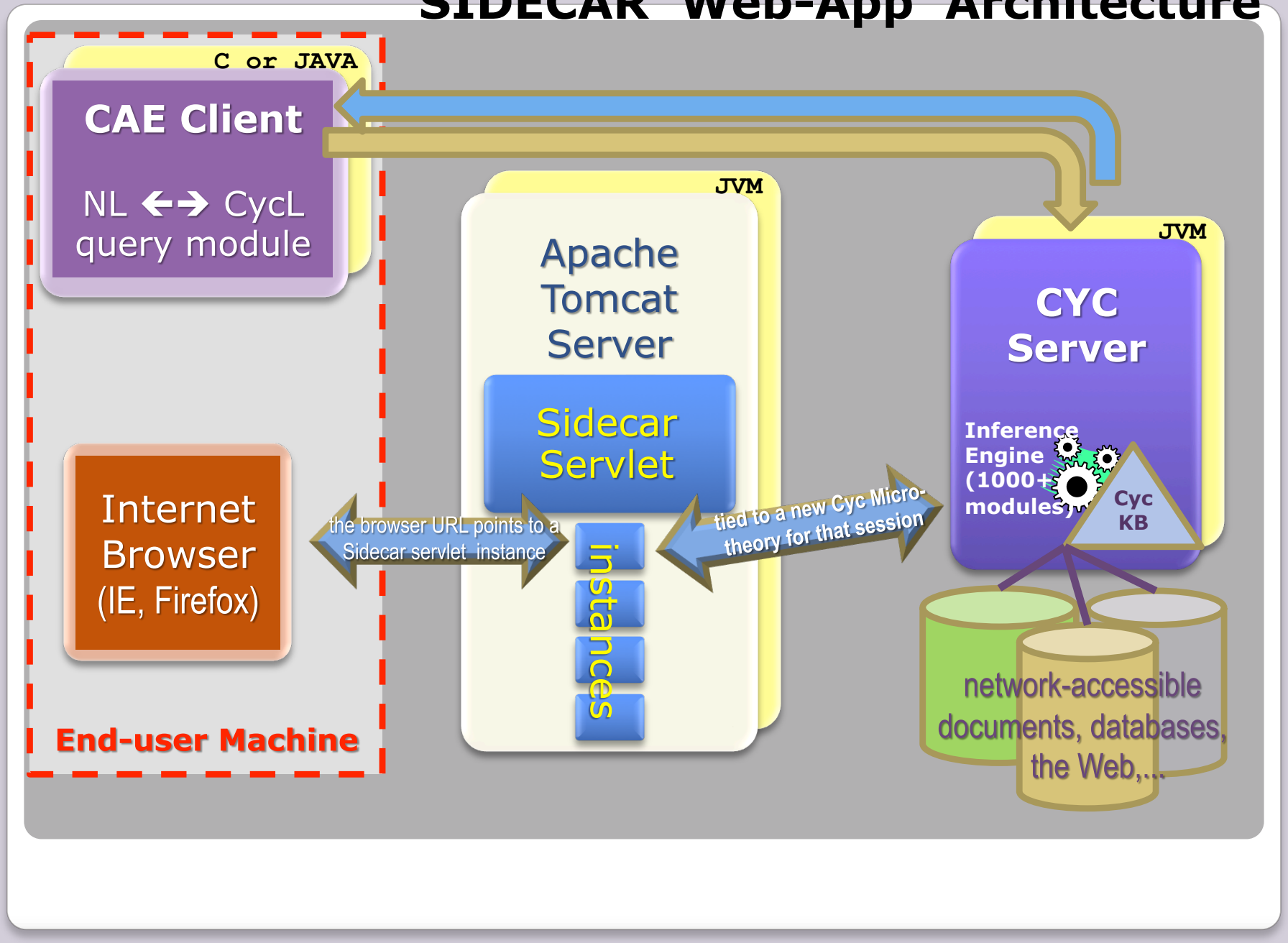
– Sometimes it induces what the most information-theoretically useful place would be, for the user to focus their attention and work on next

- **Suggest to the user where they might want to attend next (highlight or GUIDE)**

– Sometimes it induces plausible guesses for as-yet-unstated relations

- **Ask the user to confirm or deny that hypothesis (or just skip doing either)**

SIDECAR Web-App Architecture

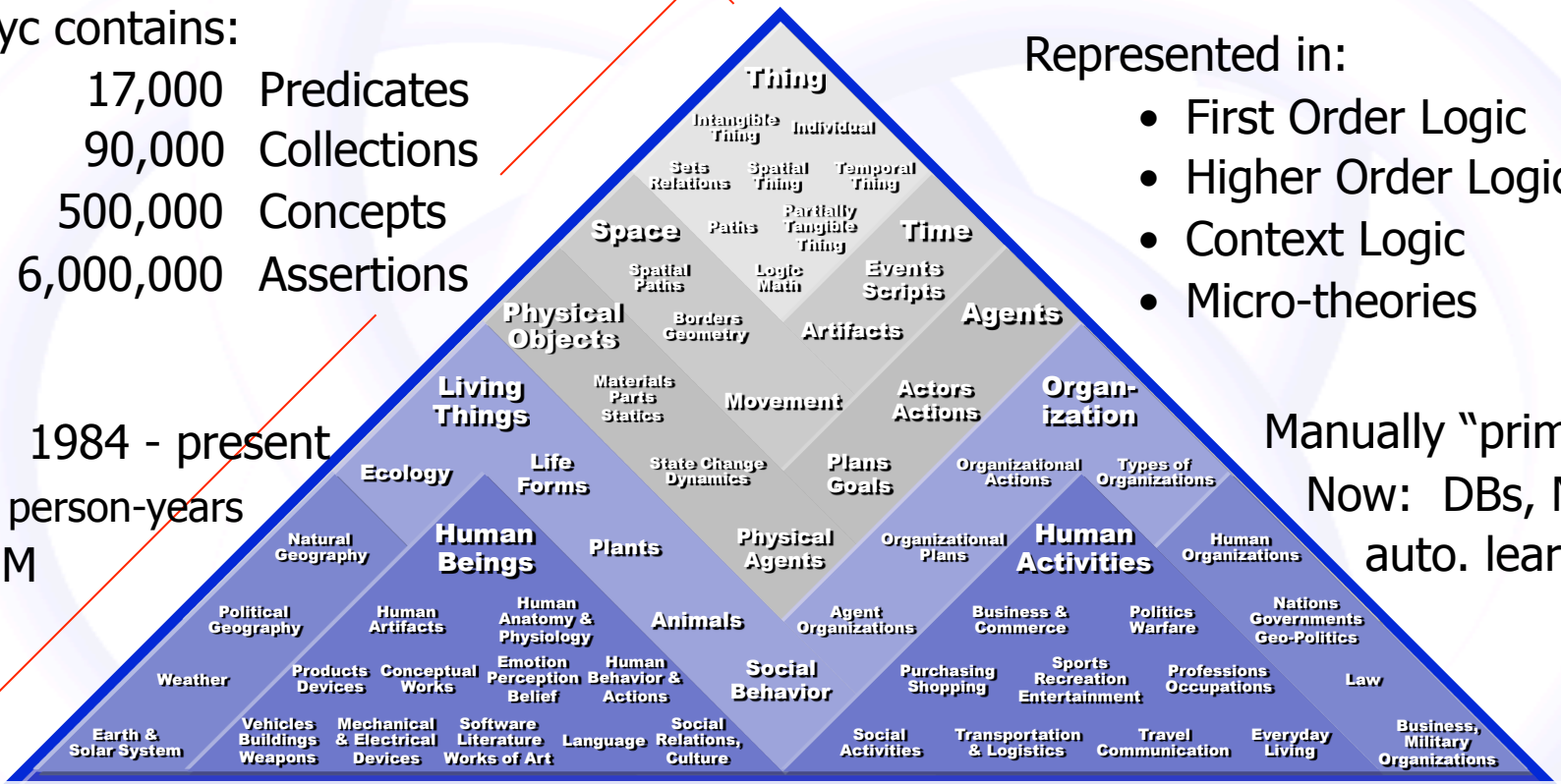


Cyc Knowledge Base

Cyc contains:

17,000	Predicates
90,000	Collections
500,000	Concepts
6,000,000	Assertions

Built: 1984 - present
1050 person-years
$110 M



Represented in:

- First Order Logic
- Higher Order Logic
- Context Logic
- Micro-theories

Manually "primed."
Now: DBs, NLU, auto. learning

Domain-specific Knowledge

Databases of detailed facts, observations, etc.

How ***formalized knowledge*** + inference helps search

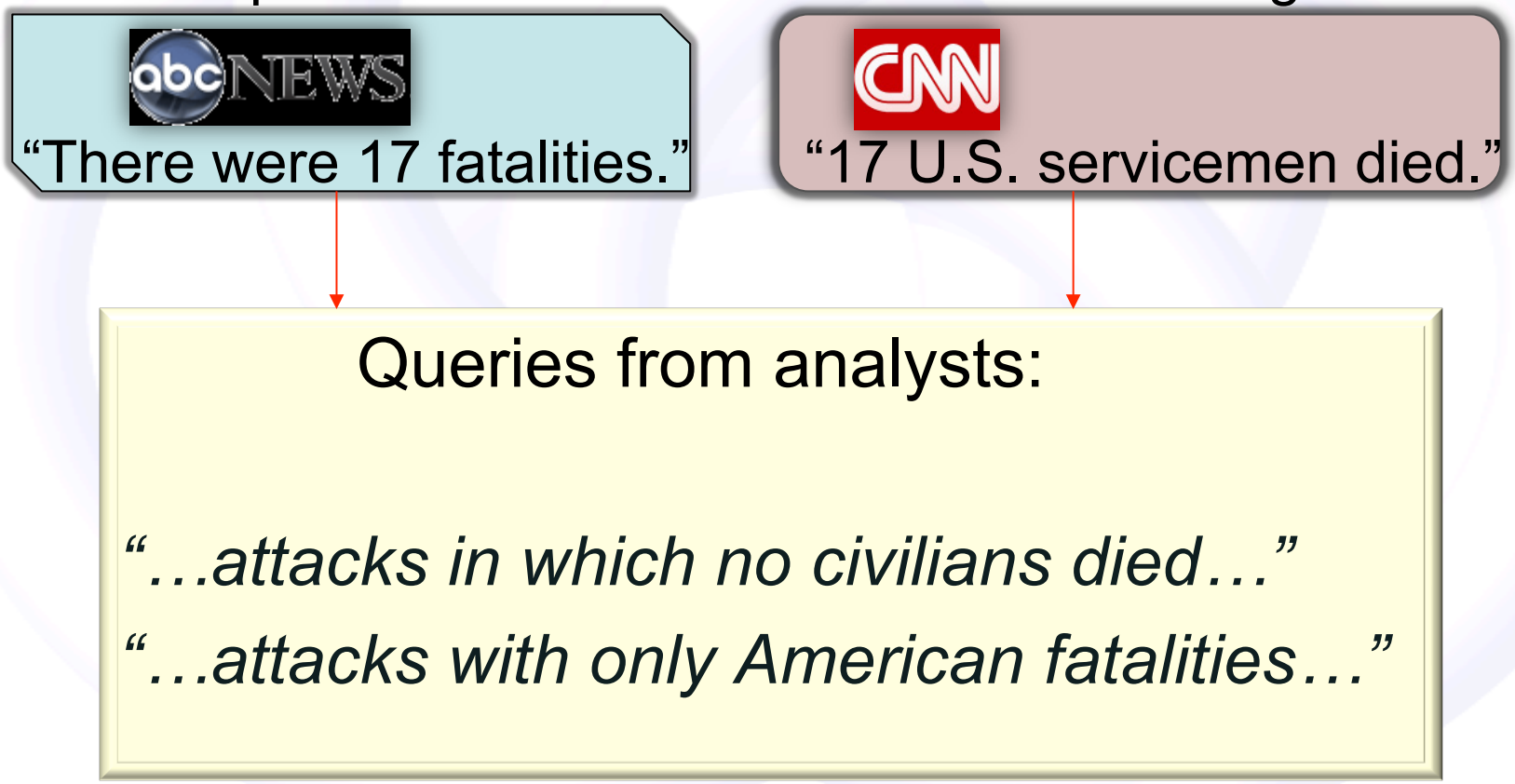
find information by inference (+KB)

Analyst's Query:

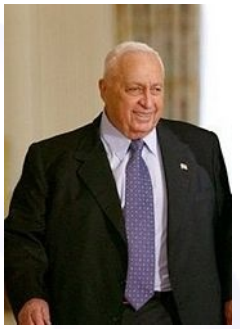
"Government buildings damaged in terrorist events in Beirut between 1990 and 2001"

How ***formalized knowledge*** + inference helps answer queries

2 reports about the Khobar Towers bombing:



Cyc Knowledge Base + Cyc Inference Engine



"Ariel Sharon was in Jerusalem throughout 2005
(except for isolated trips each < 1 week long)"

temporal reasoning

"Condoleezza Rice made a ten-day trip to Jerusalem in February of 2005"

Both of them were in Jerusalem during February 2005
(at least for a few contiguous days during that month)

diplomacy reasoning

They had a meeting, then (even if it didn't make the news.)

400+ Relations Between an Event and its Participants

- performedBy
- causes-EventEvent
- objectPlaced
- objectOfStateChange
- outputsCreated
- inputsDestroyed
- assistingAgent
- beneficiary
- fromLocation
- toLocation
- deviceUsed
- driverActor
- damages
- vehicle
- providerOfMotiveForce
- transportees

cYcorp

Next steps for development and deployment

1. Flesh out this prototype to the point where it can generate an entire ICD or CDD (e.g., PISR)
2. "Backfill" by running SIDECAR through the repository of relevant existing documents and existing policies/requirements, entering them as though they were new ← incr. interoperability
3. Augment the knowledge base so that SIDECAR can support software architects tasked with <u>designing</u> new target systems, not just <u>documenting</u> their designs
4. Continue augmenting SIDECAR's KB to the point where it can be deployed operationally
   * Collaborative effort with Faccina Global to integrate logic-based SIDECAR + English-text-based CDTM
5. Build an *ad hoc* query formulation/answering interface
   * After 2 hours of training, capabilities officers should be able to use it to generate a new requirement, to find existing similar requirements, to determine applicable policies, and to see the impact of policy changes on the new requirement.
   * A capabilities officer can support a decision-maker or analyst by entering a potential change in policy, technology, exigent threat, etc., in "what-if?" mode, to automatically calculate the impact of that on the terrain of JCIDS programs
6. Carry out periodic "regression tests" on all active programs' ICDs, CCDs, etc., re-running them through SIDECAR as though for the first time, to catch new "warnings" worth calling human attention to (due, in turn, to personnel changes, technology changes, operation priority changes, etc.)
7. Broaden the coverage (beyond JCIDS) and deployment (to other Services) and develop DoD internal capability to do the maintenance/extensions/training (rely on Cycorp only for 3rd echelon of maint.)